

C. H. Stapper
P. P. Castrucci
R. A. Maeder
W. E. Rowe
R. A. Verhelst

Evolution and Accomplishments of VLSI Yield Management at IBM

The methods developed at IBM to manage and improve the yield of some of its newer FET semiconductor products are described. A number of visual inspection and electric monitoring techniques have evolved since discrete semiconductors were manufactured. The data obtained with these techniques are used in self-checking yield models to give the relative yields for all the yield components. The results are applied not only to day-to-day control of the manufacturing lines, but also in the long-range forecasting and planning of future semiconductor integrated circuit products. An example is given comparing the actual and planned yield of a 64K-bit random access memory chip as a function of time. The results show the yield enhancement that was obtained with redundant circuits and additionally with the use of partially functional products. Another example shows the decrease in fault levels over a span of more than ten years.

Introduction

The manufacture of semiconductor components using very-large-scale integration (VLSI) technology typically requires that literally millions of microscopically small components and their interconnections be fabricated on the surface of a single silicon wafer. An example of a logic product chip is shown in Fig. 1. It is usual for many identical logic or memory chips to be fabricated simultaneously on the same wafer. Manufacturing is accomplished in a series of sequentially applied processing steps that takes about three months from start to finish.

Because of the complexity of the fabrication process and the microscopic dimensions involved, defects, such as flaws in the material structure or dust particles that settle on the wafer, or processing errors, inevitably prevent some portion of the chips from operating as desired. That portion of testable chips on a wafer which do finally operate correctly is called the manufacturing final test *yield*. Yield is a fundamental measurement used to plan and control the manufacture of semiconductor components. It is closely related to the

chip size, the density of contamination on a wafer, and the resulting defects in a chip. Increasing the yield essentially means learning how to reduce the number of defects per unit area.

At IBM, the management and technical system that has evolved in an effort to maximize the yield of semiconductor products is a unique application of process feedback loops that make use of self-correcting statistical yield models. Use of this system permits process engineers not only to make process corrections while the wafer is only part way through the fabrication sequence, but also to determine expected yields for future products with increased component densities, to determine manufacturing resources needed for products being prepared to enter production, to estimate manufacturing volumes that can be achieved, and to project the rate of yield learning that will be needed over the life of a product. The system is vital to both the long- and short-term management of semiconductor manufacturing operations.

Experience during the past fifteen years has shown that

© Copyright 1982 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

the trend towards higher and higher component densities has continually stimulated the development of improved defect control techniques, and that the achievement of better control has, in turn, paved the way for introducing the next level of component integration. This paper traces the evolution of IBM's yield management system, discussing some of the key contributions made along the way, and attempts to illuminate the problem-solving approach that has so far proven very productive.

Origins of yield management in semiconductor manufacturing

Evolution of the IBM yield management system to its current state of development has come about through a continuing series of efforts to improve tools for detecting defects at various stages of manufacture and to improve techniques for analyzing the effects of each type of defect encountered. The system now in use has its origins in the approach to defect detection and analysis used by IBM in the mid-1960s in manufacturing its first semiconductor components—the Solid Logic Technology (SLT) family of computer logic circuits developed for the IBM System/360, announced in April 1964 [1].

In SLT circuits, transistors and diodes were discrete components mounted on substrates that contained active and passive components. Each transistor was made on a single die. Yield control in the production of these components consisted primarily of inspecting the photolithography masks to determine if any parts of the image were missing. Yield analysis was done by investigating products that had failed final test to identify the cause of the failure. This information was then fed back to technicians on the manufacturing line so that they could alter the processes in an attempt to improve the yield.

A flow diagram representing the SLT-era yield control method is shown in Fig. 2(a). Because major testing occurred only after the device was completely fabricated, reaction to process errors was very slow. No attempt could be made to eliminate defect-producing mechanisms early in the process sequence.

Some improvement in this yield control system occurred with the introduction of tests at intermediate points in the SLT fabrication process to measure transistor gain and collector resistance. However, there was no compelling motivation to develop methods that would allow salvaging partially completed SLT chips, since—with just one device per chip—it was economical to keep only those that passed final test, and to scrap the remainder.

It was the extension, in the late 1960s, of the integrated circuit concept to include multi-transistor memory and logic

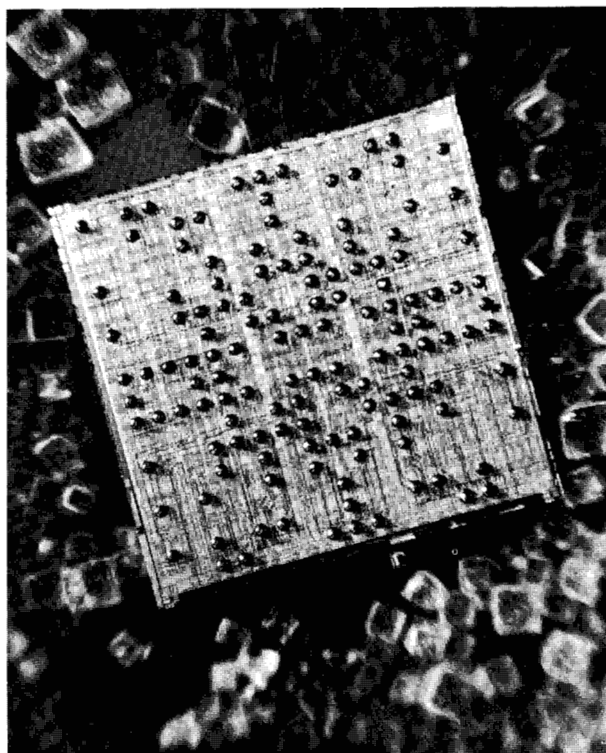


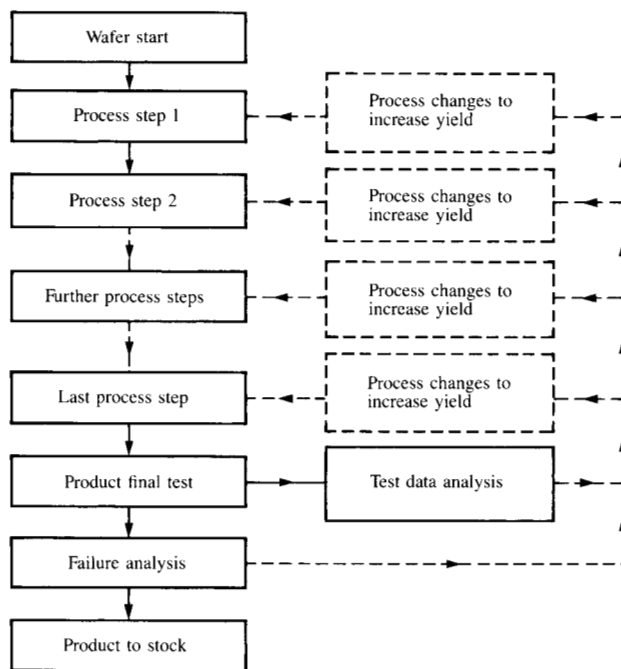
Figure 1 East Fishkill logic product. The small spheres are lead-tin balls that will connect to strips of metal wiring on ceramic substrates with IBM's flip-chip mounting technique.

chips that provided the stimulus for development of in-line wafer testing and defect analysis. Simultaneous fabrication of multiple circuits on the same wafer increased the value of wafers in process and placed a heavy penalty on having to discard good circuits along with the defective ones. Hence, real-time response to the discovery of defects became, for the first time, an urgent goal of the semiconductor manufacturing effort.

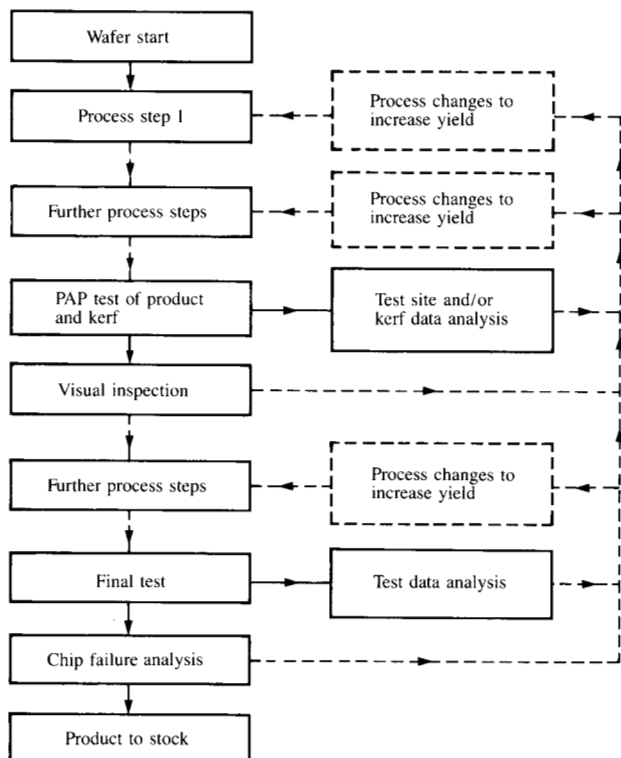
Defect analysis using post-aluminum probe testing

One of the earliest efforts to improve the yield diagnostics for integrated circuit manufacturing was the introduction of a test point that allowed the circuits to be evaluated before they reached final test. This test point, called post-aluminum probe (PAP), occurred at an intermediate stage of processing, after the electronic elements on the wafer had been completely fabricated and the first layer of metal interconnecting them had been deposited.

Measurements made at PAP became the center for yield diagnosis activity on the manufacturing lines. A flow chart showing the new test point in the fabrication process is given



(a)



(b)

Figure 2 (a) Manufacturing flow diagram for discrete semiconductor components. Failure analysis data and test data analysis can activate the feedback loops shown with dashed lines. These feedback loops were too time-consuming for LSI manufacture. (b) A test point added to the mid-process increased the speed of the feedback loops and added relevant data from test devices in the kerf, or test sites on the wafer.

in Fig. 2(b). Initially, data for PAP testing were obtained from special test devices fabricated in the product kerf, the area between the chips that is removed by the saw blade that dices the wafer. Putting test devices in the kerf, of course, conserves the usable space on the wafer for product use.

The PAP tests measure the values of certain electrical parameters of circuit elements. For example, they measure transistor gain, conductor resistivity, and contact resistance of bipolar transistors, the earliest form of integrated circuit transistors. When field effect transistors were introduced, PAP tests measured their threshold voltages, transconductances, and electrical gate dimensions.

Data collected from PAP testing of kerf devices showed that not only did parameter values vary from wafer to wafer, but significant variation in data taken from different locations on the same wafer could occur. To correlate these variations with the functional operation of the product, it was necessary to develop tests that could be applied at PAP. With the use of such tests, it became possible to map particular categories of failures of the functional test results and to compare these with maps of the parameter values.

In this way, it became possible to determine the complete *process window*; that is, to specify functional yield as a function of parameter values for a given circuit design. This process window information was used to center and maintain the process for optimum functional yield. Process window data, as well as parameter distributions, were also fed back to circuit designers to help them make design improvements and to help ensure that future designs could make optimum use of tradeoffs among parameter values, yield, and circuit density.

The extension of integrated circuit technology to the fabrication of memory cell arrays stimulated an important advance in the analysis of yield data obtained at PAP. For the first time, it became possible for IBM engineers to locate defects easily and to quantify the effect of each type of defect on yield. The key element of this new approach to analysis was the *bit failure map*. This map, a computer printout, showed the locations of memory cells that had failed functional testing within a chip at PAP. Subsequent visual inspection of the wafer revealed certain defects which could be directly related to information on the bit failure map. An example of a bit failure map is shown elsewhere in this issue [2].

As a result of information gained through PAP defect analysis, it became clear that most of the defects responsible for functional test failures were introduced during the photolithography steps and the metallization processes. It also became apparent that the silicon material itself was a source

of yield loss. One type of material defect called a *pipe* was found to result from enhanced emitter diffusions along crystalline dislocation lines or stacking faults in the base region of bipolar transistors [3]. Pipes, which cannot be observed with conventional optical microscopes, caused electrical shorts between the emitter and collector regions of the transistors.

In order to better understand and quantify this yield variable, D. P. Martin of IBM Burlington conceived the idea of developing a special mask to permit the use of product chips as a *test site*. Martin's test site is made on standard product wafers that have been fabricated to the stage at which contact holes are etched into the product. At that point, special metallization patterns are applied. One of these patterns joins all the emitters on a chip, while another pattern joins all the collectors (see Fig. 3). In this way, the sensitive or critical area for pipe defects is increased, and the metal patterns can be tested to estimate the number of chips on the wafer that have pipe defects. If desired, after the pipe yield data have been collected, the special metallization patterns can be chemically removed, and the test wafers returned to the manufacturing line for the application of the standard product metallization patterns. Thus, the use of test sites can provide useful data for yield control without any sacrifice in the amount of product being made.

Leakage defects are another type of failure mechanism that cannot be detected optically. With the use of scanning electron microscopy, H. Boisselle discovered that tiny (submicron) phosphorus *spots* in the oxide layer of field effect transistor gates could cause charge leakage within memory cells. H. J. Geipel and W. K. Tice [4] used bit failure maps collected at PAP to isolate and understand the behavior of leakage defects, and were able to develop methods for increasing leakage defect yields in dynamic FET memories. To measure the leakage defect densities and determine the magnitude of the yield losses, test sites were needed. During the early 1970s, a number of different leakage defect monitors were crucial for evaluating the experiments that resulted in the reduction of leakage defects and opened the way for the volume manufacture of dynamic FET memory chips. Measurements on such test sites are still made daily to ensure defect control in manufacturing.

Pinholes in dielectric insulation layers are another defect type that is difficult to detect with optical microscopes. Test structures for these defects usually consist of simple metal or polysilicon patterns on top of the insulators. In the development and manufacture of FET technologies, such structures were and are still being used in the kerfs of product wafers and on monitor wafers. In this way, the integrity of FET device gate oxides and pinhole defect density levels are monitored continually.

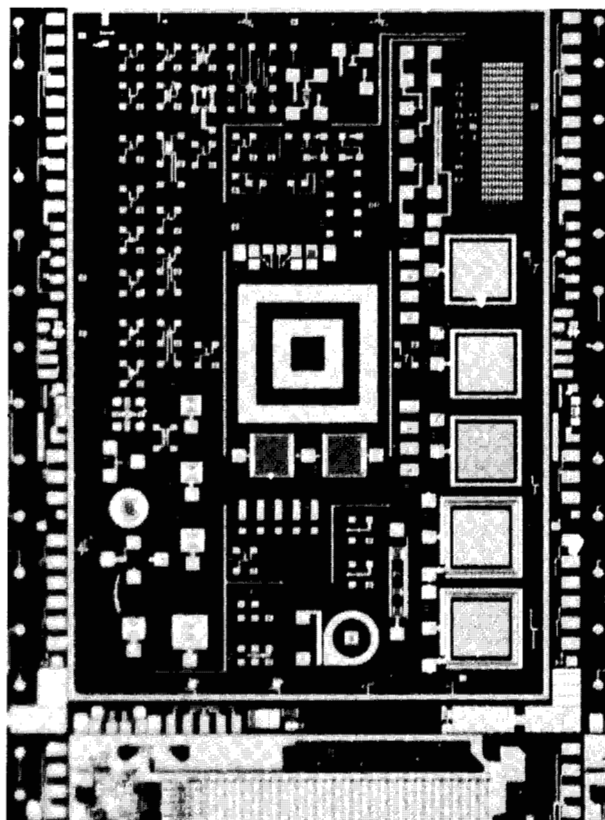


Figure 3 Typical test site. Emitters and collectors are connected with a special metallization pattern for finding emitter and collector short circuits caused by "pipes."

D. R. Thomas of IBM Burlington and A. V. Satya of IBM East Fishkill introduced photolithographic defect monitors to study the effect of smaller designs or ground rules. Such monitors, which consist of conductive serpentine lines and interspersed fingers, have also been used elsewhere in the industry [5]. Electrical monitors were also developed to measure line width and pattern alignment [6]. Many of these concepts have been incorporated as structures in the product kerf at a number of IBM locations.

In-line controls

The PAP analysis method, including the use of kerf devices, functional tests, bit failure maps, and chip failure analysis, evolved during the late 1960s. Since that time it has come into use at all of IBM's integrated circuit manufacturing locations and even now continues to be a central element in the company's yield and defect control system. But despite its usefulness, PAP analysis still requires the completion of a substantial number of processing steps before it can be fed back to make process corrections. Used by itself, PAP analysis results in slow yield learning and leaves open the



Figure 4 Bright-light inspection station used for measuring particulate contamination on wafers.

possibility that undetected catastrophes, occurring early in the process, will destroy a considerable amount of the product. Thus, in order to forestall such disasters, some form of *early warning* scheme has been inserted in the manufacturing process since the earliest days of semiconductor production.

In the manufacture of discrete bipolar transistors used in the SLT technology, engineers learned to include *control wafers* along with regular production wafers at certain critical steps in the fabrication sequence. These wafers, usually left blank, were processed only at the diffusion or oxidation steps. Afterward, their resistivities and oxide thicknesses were measured and, if the measurements were out of specification, process errors leading to the incorrect values were immediately corrected. SLT engineers also used samples of control wafers to determine the depths of transistor junctions and to infer impurity profiles from resistance measurements made after successive oxidation and etching steps.

Measurements of control wafer characteristics continued to be made when semiconductor technology moved into the manufacture of multitransistor integrated circuits. This technique proved to be very effective in preventing process disasters. It is, however, very difficult to correlate control wafer data with final test results on completed chips. This is because control wafer data represent only a small sample of the parameter measurement distribution found in the actual product wafers. Only by using very large samples of control wafer data, obtained over an extended period of time, is it possible to correlate these data with parameter distributions obtained at PAP tests.

D. P. Kennedy, P. C. Murley, and R. R. O'Brien [7] developed a Monte Carlo technique that could explain the

distribution of values of discrete-transistor gain on the basis of control wafer data and in-line temperature measurements. This technique proved to be valuable in the manufacture of early bipolar-transistor integrated circuits since it could show which process variables had the most influence on the distributions of device parameter values. The method was extended to field effect transistors on a static memory chip by C. H. Stapper and P. B. Hwang [8].

Other techniques besides the use of control wafers were developed in an effort to obtain in-line defect data. Among these were two kinds of visual inspection methods for identifying random defects in photolithography masks and in the wafers being processed. One of the methods, called *bright-light* inspection, was developed to detect the presence of foreign materials or pattern irregularities by scattering a beam of light incident at a small angle to the surface of a mask or wafer (Fig. 4). Particles less than a micrometer in diameter can be detected in this fashion. Bright-light inspection allowed engineers to gain an insight into the control of cleanliness that had not been available previously. Standardization of measurement has allowed comparison of cleanliness levels among the many facilities within the company. As a result, significant improvements have been implemented in handling procedures and manufacturing contamination control, both necessary to achieve low defect results.

Although the particulate control obtained with bright-light inspection led to increased cleanliness, it was difficult to determine its effect on yield. The photo inspections that were introduced in IBM's integrated circuit manufacturing lines during the late 1960s were a more direct method for getting to the heart of the yield problem. For the first time, an attempt was made to obtain an in-line count of the same defects that were causing failures at PAP, functional test, and final test.

Photo inspectors were instructed to look for defects through microscopes and to determine whether these defects would cause failures at final test. This type of inspection was performed after each photo step, as shown in the flow diagram in Fig. 5. The feedback from each inspection to its preceding photo process is very short. Also shown in Fig. 5 is a feedback loop for the failure criteria obtained from final test chip failure analysis. It was this feedback loop and the one from chip failure analysis at PAP, that became the weak points of the system. As the 1960s drew to a close, it appeared that as much time was being spent understanding the information as it took gathering it. Continued redefinition of faults as a result of changing failure analysis results, as well as the uncertainty of knowing which process sector caused the defects, made this system difficult to manage. The solution to this problem was found by introducing yield models into the control system.

Development of yield models

Yield models are essentially mathematical relationships that attempt to predict the effects on product yield of various defects encountered in the manufacturing process. An early yield model relevant to semiconductor manufacturing was developed in 1964 by T. J. Lawson, Jr., of IBM Owego [9]. N. Haddad of IBM East Fishkill also developed a simple model based on Poisson statistics which was applied to the manufacture of 64-bit and 128-bit memory chips made by IBM in the mid-1960s. These models did not distinguish between technologies. What was needed was a model that not only contained the most apparent yield detractors found at PAP, but would also give relatively realistic comparisons between different products. Such a model was developed by R. H. Dennard of the IBM Thomas J. Watson Research Center [10]. Dennard used his model originally to project the difference in yield between bipolar and FET static memory chips. Since that time, models of this type have been used for much of the planning of LSI and VLSI products at IBM.

In his model, Dennard accounted for effects of random defects by categorizing them into different types and then calculating a yield resulting from each type. He also took size into account. It was determined experimentally that, for the defects affecting the photo patterns, the defect size distribution varied approximately as $1/x^3$, where x is the defect size. An example of such a size distribution with some recent data is shown in Fig. 6. Dennard also developed a method for determining the sensitivity of photo patterns to defects of different sizes.

Dennard's methods were subsequently developed into computer models by R. W. Bartoldus and N. F. Brickman of IBM East Fishkill, and by W. N. Kuschel and D. H. Withers of IBM Burlington. Combining the sensitivity to defects with the defect size distribution results in what is called the critical area or effective susceptible area for the photo patterns. The yield for each defect type can be estimated [9-12] by

$$Y_i = \exp(-A_{ci}D_i), \quad (1)$$

where D_i is the average defect density of the i th defect type and A_{ci} is the critical area of the i th defect type. The total yield for the product, including N different random defect types, can then be written as

$$Y_t = \exp\left(-\sum_i^N A_{ci}D_i\right). \quad (2)$$

Non-random effects such as misalignment of photo masks, over- or under-etching, and out-of-specification electrical parameters also contribute to yield loss and must be accounted for in the yield model. As long as all such yield detractors are independent, their individual yields can be multiplied together to give a total gross yield, Y_0 , which

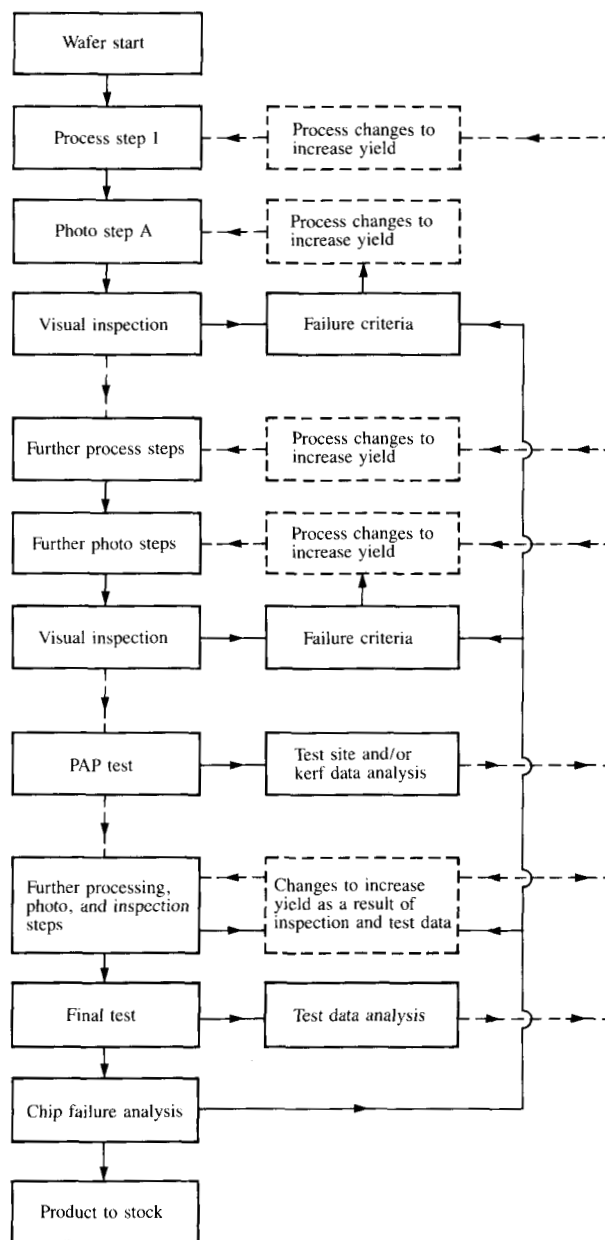


Figure 5 Introduction of visual inspection after photo steps. The feedback loops were unstable due to changing failure criteria.

enters the model as

$$Y = Y_0 \exp\left(-\sum_i^N A_{ci}D_i\right). \quad (3)$$

This is the expression that became the basis for projecting the final test yield in IBM semiconductor manufacturing. Some modifications of this formula, due to the use of different statistics, are given in Appendix A.

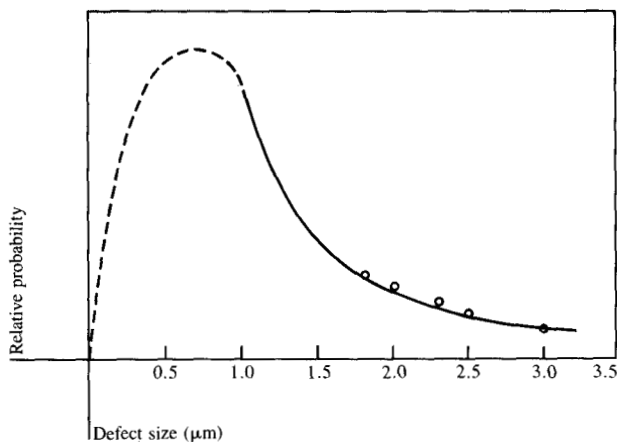


Figure 6 Defect size distribution. Data points are measured with electrical defect monitors in the SAMOS manufacturing line. The $1/x^3$ curve gives the best fit to the data with $x > 1$. The smallest monitor-measured defects are $1.8 \mu\text{m}$. It is assumed that the size distribution will decrease for smaller defects, hence the dashed curve in the region $x < 1$.

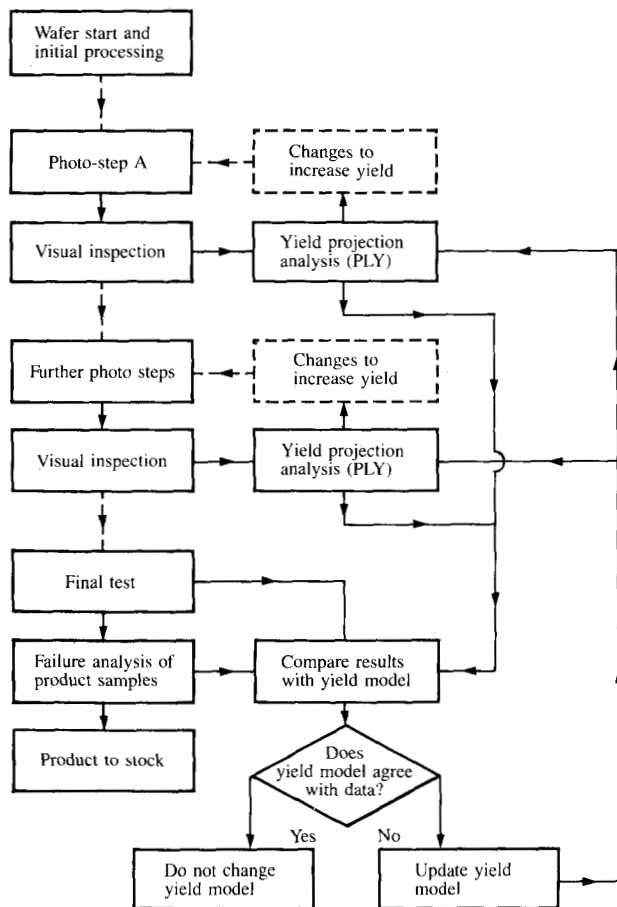


Figure 7 Visual defect data from in-line inspection are analyzed with a statistical yield model to obtain the photo-limited yield (PLY) for each photolithographic step. The statistical models are checked against the final test yield and updated, if necessary. For simplicity, other feedback loops from post-aluminum probe (PAP) data are not shown here.

The discussion in this paper has centered on the control of yield as measured at final test. Another type of yield also plays a role in the total productivity of the manufacturing line. This is called *process yield*, which is the percentage of started wafers processed to completion. Whereas final test yield is exclusively a function of defects, process yield losses, by contrast, are caused by mishandling, breakage, equipment malfunction, and process errors. Wafers lost in these ways are usually scrapped at the point where the error or problem is detected, and never reach final test. Multiplying process yield by final test yield gives a total measure of manufacturing productivity. Process yield is further discussed in Appendix B.

Yield control by model

The key to effective in-line control of product yield was an observation by three Burlington engineers that makes it possible to predict that certain visual defects will cause specific failures at final test. R. A. Maeder, F. W. Oster, and R. J. Soderman reasoned [11] that, even though a certain type of defect might not cause a failure at the location where it is observed, this same defect could cause a failure if it occurred in a more sensitive area of the chip. There exists, therefore, a certain probability that a defect of a given type will cause a failure on the chip. Yield can thus be determined by counting the average number of defects per chip and entering the result in a yield model like that of Eq. (3). Because the defect data are taken from inspections made after each step in the photolithography process, the expression is called a *photo-limited yield (PLY)* model.

The flow diagram for a control system that makes use of the PLY inspection model is shown in Fig. 7. The accuracy of the model is continually checked and improved by comparing the calculated yields with the actual yields determined at final test. If the two values are different, chip failure analysis is used to pinpoint the reason, and the probabilities of failure attributed to the recalcitrant defect types are revised. The system is therefore self-correcting.

Like the results with test devices in the kerf, the PLY data showed strong wafer-to-wafer and within-the-wafer variations in defect counts and yield. In fact, the yield appeared to vary radially over the wafers—an effect that had also been reported by Yanagawa [12]. As a result, it was decided to control this effect by collecting the defect and yield data from concentric regions. The PLY data gave IBM manufacturing engineers the capability of measuring the effect of process changes for the first time. The PLY model for defect monitoring has made it possible for process engineers to quantify their own process sector's influence on the final test yield. They no longer have to wait months to get actual final test results to ensure that the changes they made worked. It allows engineers to divide the product yield into its constitu-

ent components, enabling them to use yield models as a sort of bookkeeping system. Missing yield components can be readily detected by comparing the calculated yield to the actual yield. Furthermore, when problems arise, management can use the system to establish priorities for manpower allocation.

As a result of the PLY inspections, some yield detractors have been completely eliminated. For example, PLY data are used to remove photo masks from the line when the mask-limited yield becomes too low. Today, this has been developed to such a degree that some IBM manufacturing lines will not accept a mask until data are available to prove that it is free from defects.

The yield modeling concept was also extended to defects that could only be detected by defect monitors. Although special wafers with defect test sites were processed in the early 1970s, it was not until the middle of that decade that D. R. Thomas of IBM Burlington introduced defect monitors in the product kerf. Even though the data were collected at PAP, it became possible to measure defect levels with larger samples than for PLY inspections. The results allowed more accurate statistical evaluations for the defect density distributions, and therefore led to improved yield models [13, 14]. The kerf monitors were also used for the measurement of leakage and oxide pinhole defects that were not detected with PLY inspections and could only be found before that time by means of very tedious and costly delayering techniques. The monitor data became a superb tool for measuring the process changes that could be made to increase the leakage and pinhole yields of the high-density FET memory chips. Miniaturized versions of these defect monitors were designed by P. Hyde of IBM Burlington for the manufacture of IBM's 64K-bit chip. These monitors were of key importance for tuning new manufacturing tools and for decreasing the defect levels that were attributable to them.

Defect reduction trends

The effort in IBM to develop increasingly better defect control techniques has produced a downward trend in defect levels for both bipolar and FET memory chips. PLY inspection (see Fig. 8 for a typical station) was introduced in 1970. From that time until 1973, as shown in Fig. 9, the fault densities decrease sharply, proving the effectiveness of the technique. (A fault is defined as a defect that produces an electrical failure.) The period from 1973 until 1976 indicates a more gradual rate of learning. Then 1977 shows a new burst of fault density learning resulting from better contamination control, new photo tools, higher mask quality, and improved inspection techniques for high-yield wafers. The dashed line, from 1979 on, is a projection toward the defect levels that must be reached in order to produce megabit memory chips and computer CPU chips.

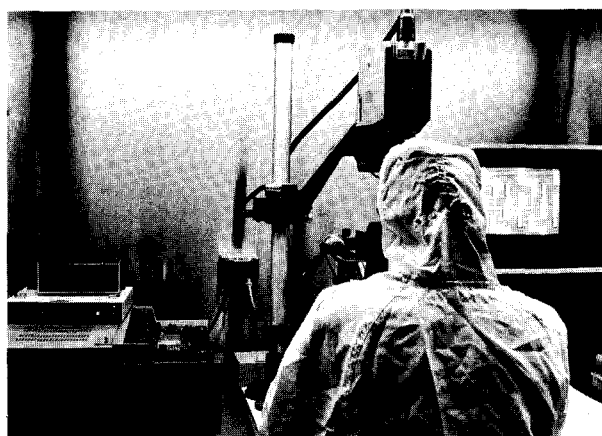


Figure 8 Photo-limited yield (PLY) inspection station. Video pictures are recorded on video tape. The inspector's voice-comments on what was observed are also saved.

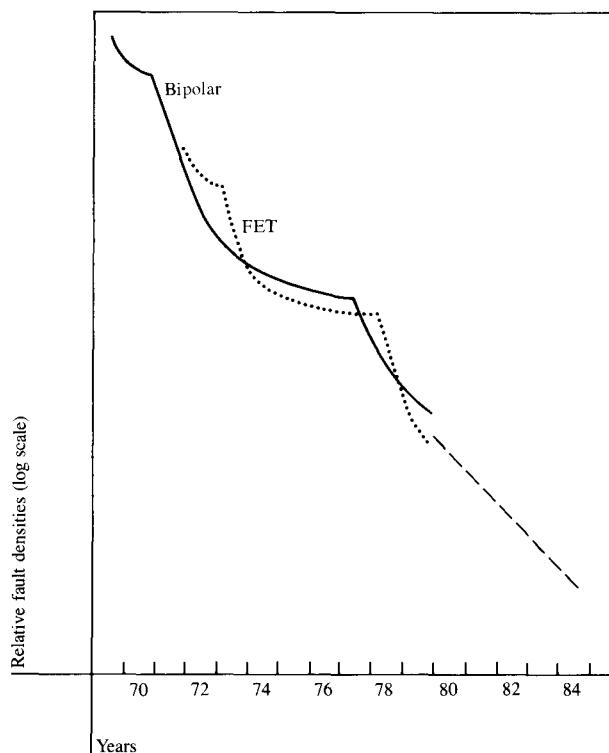


Figure 9 Fault level trend for IBM bipolar and FET memory products. The dashed line beyond 1979 shows the improvements required to make megabit memory chips and full CPU chips a reality.

At low defect density levels, two phenomena occur that affect the accuracy of defect density measurements. One is purely statistical and occurs for PLY and for defect monitor

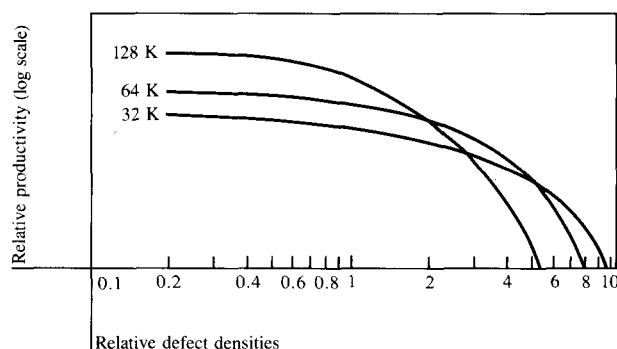


Figure 10 Productivity in bits per wafer as a function of defect density levels for random access memory chips of different bit densities.

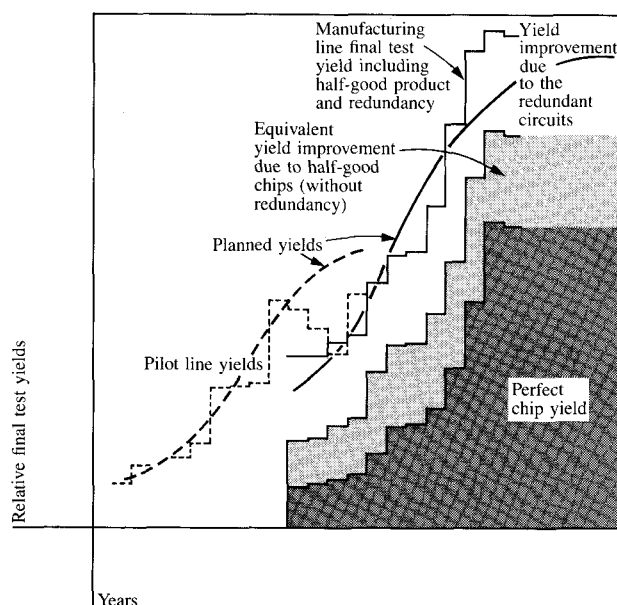


Figure 11 Yield plot of IBM's 64K-bit random access memory (RAM) chip. The gain in yield due to redundant circuits and to the use of partially functioning chips is shown in relationship to the yield of perfect chips for the manufacturing line data.

data. For example, if a PLY inspector finds 15 defects in a given sample, the 95% confidence interval is between 8.4 and 24.8 defects. This implies a statistical uncertainty between +65% and -56%. If, however, the inspector finds only two defects in the same sample, the 95% confidence interval is between 0.2 and 7.2 defects, or +260% and -90%. Thus, the fewer the defects found in a sample, the greater will be the range of statistical uncertainty in the defect density. While this effect can be easily reduced by increasing the sample size, there is, however, a better and more productive way to ensure accurate defect density estimates: begin production of a new product.

This solution also works for the second phenomenon that affects the accuracy of defect measurements. It has been found that, when there are fewer defects on a chip, it becomes increasingly difficult for psychological reasons for the PLY inspectors to find them. (This is because *many more* samples are required to obtain the same confidence limits.) Hence, when a product reaches a certain yield level, it usually becomes more productive to start fabricating a denser chip. An example of this is given in the graph in Fig. 10, which shows how the number of megabits per wafer for three different types of memory chip varies as a function of defect density. It can be seen that as the defect densities decrease, the chips containing more bits become more productive than chips with a smaller number of bits. Since the larger and denser chips will have more defects, the problem of having too few defects to detect has been neatly eliminated.

Visual inspection also becomes less accurate as the chip complexity increases. In the manufacture of the IBM 64K-bit memory chip, it has been observed that the PLY inspectors find between 60% and 100% of the defects that cause failures. The average is better than 90%. The 60% accuracy occurred only on the first metallization layer, and it still appeared perfectly adequate for controlling the line. Defect reduction on this very complex metallization process has proceeded rapidly. One can conclude, therefore, that much more complex chips can be manufactured before the PLY technique becomes obsolete.

Some integrated circuit manufacturing lines at IBM have been making the same chips for an extended period of time. Yields in these lines tend to saturate between 80% and 90%. Manufacturing engineers are well aware of the limitations in PLY accuracy for such products. In one such case, to prevent large inspection samples and high inspection costs, a system called TPLY (tested PLY) was developed that was a hybrid between the control schemes shown in Figs. 5 and 7. To help the PLY inspectors find defects, sample lots are processed through the line at high speed and tested at a test point. Both the wafers and the bit failure maps that result from the test are then given to the PLY inspectors for counting the different defect types. Although this method may resemble feedback from chip failure analysis, it is not the same. Failure analysis requires a high degree of expertise and is destructive, whereas inspection by the PLY inspectors is nondestructive. Furthermore, by having a road map to the first few defects, the PLY inspector will discover more defects if there are any. This hybrid method allows manufacturing engineering the freedom to shift into either a PLY or TPLY mode when the data require it. For very-high-yield products, TPLY has proven to be an adequate control.

Redundancy

The use of redundancy in integrated circuit memories is one way to increase manufacturing yield without attempting to

alter the fabrication process. Extra (redundant) memory cells can be fabricated on the chip and used to replace defective cells. S. E. Schuster (IBM Yorktown) has calculated the memory chip productivity gains expected from use of redundant word and bit lines [15]. N. Brickman, W. F. Mikhail, S. P. Bennett, and W. E. Donath (IBM East Fishkill) and A. N. McLaren (IBM Burlington) have developed computer models for designing memory chips with redundancy. The use of redundancy to increase the yield of 64K-bit memory chips has been reported at the Bell Labs [16-18]. More recently, Intel [19, 20], Inmos [21], Mostek [22], and Hitachi [23] have used redundancy in chip design.

The yield model described in [14] was used in projecting the yield for IBM's 64K-bit memory chip, which includes over 1000 redundant cells. Examples of the projected and actual manufacturing yields for this product are shown in Fig. 11. The results shown are for two process lines. The earlier yields are from a pilot line which made the engineering hardware for prototype computers. A special manufacturing line was built for the fabrication of memory chips (sample shown in Fig. 12) used in the computers intended for shipment to customers. Also shown in Fig. 11 are the perfect chip yields, that is, the actual yields without the benefit of redundancy and partially good product.

To make the learning possible along the projected yield curves, the model projections are also made for the individual yield components. These components for the 64K-bit memory chip are shown in Table 1. They consist of 11 photo-related type defects, 5 defect types usually caused by the process and 8 miscellaneous yield limiters, including losses caused by circuit failures. Individual targets were established for each yield detractor and plans were made to see that these targets are achieved. By focusing attention on those yield components not making their targets, manufacturing management is able to allocate the proper resources to get to the bottom of the problems and to solve them.

Also shown in Table 1 are the diagnostic techniques available for measuring the progression of individual yield detractors. It can be seen that there is a considerable overlap among these methods. The defect monitors, for example, are needed primarily to monitor junction leakage defects, pinholes, and voltage-sensitive defects [24]. The same probe and tester that measure these defects simultaneously check for the existence of open and short circuits in other conductor patterns. Chip failure analysis, furthermore, must overlap with the other measurements since its major function is to check and verify the other techniques and to ensure that no defects pass undetected.

Technology improvement interchange

The integrated circuit inspection techniques and monitoring schemes we have described have another benefit at IBM.

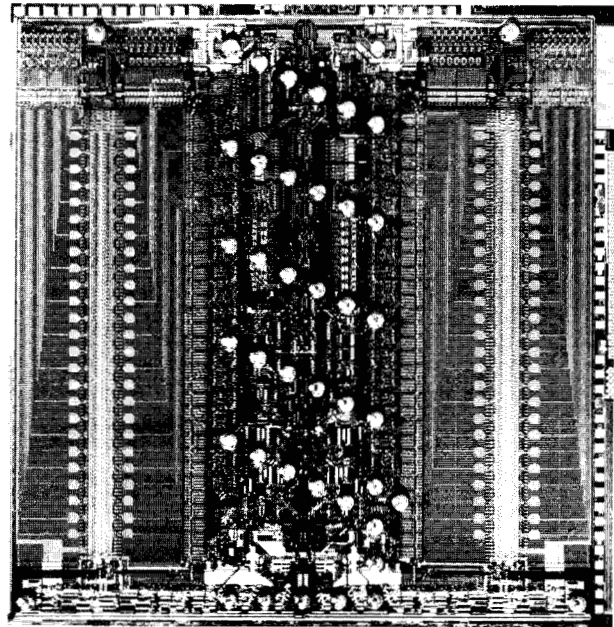


Figure 12 A 64K-bit random access memory chip. The lead-tin balls are in the center and at the corners of the chips. Fuses, which are blown to steer data to redundant word and bit lines, are at the center of the two evergreen-shaped wiring patterns located over the memory arrays.

With a number of plants in operation, there is considerable opportunity for comparing improvements. The achievement of high yield in any of the plants is immediately shared within IBM. A manufacturing innovation causing a yield breakthrough in any one location is adapted very quickly by the other manufacturing lines. Even the mundane day-to-day monitor results are scrutinized between locations, and methods of improvement are implemented. For instance, yield detractors occurring in the manufacture of the 64K-bit memory chip are reviewed periodically between the Sindelfingen plant and the Burlington plant. The capability of measuring the yield components has made it possible to minutely compare the differences between the two locations. As a consequence of these exchanges (and hard work on both sides of the ocean) the yield at both locations has increased and has resulted in consistently high computer memory productivity. In a similar way, the production of logic products is compared between Essonnes and East Fishkill, and between Manassas and Burlington. In all cases this interchange of technology experience has increased manufacturing productivity within all participating locations.

Conclusions

We have shown in this paper that the yield learning does not have to be a haphazard trial and error venture, and that it can be approached scientifically. Realistic yield goals must be set by those responsible for making them. For some defect

Table 1 Yield detractors measured for learning of IBM's 64K-bit memory chip. Functional analysis and failure analysis make use of bit failure maps.

	<i>Diagnostic techniques</i>			
	<i>PLY</i>	<i>Final test</i>		
	<i>In-line visual inspection</i>	<i>Functional data</i>	<i>Failure analysis</i>	<i>Kerf monitor data</i>
<i>Photo levels</i>				
1. Diffusions (missing)	X		X	X
2. Diffusions (extra)	X	X	X	X
3. Thick oxide	X		X	
4. Polysilicon (missing)	X		X	X
5. Polysilicon (extra)	X	X	X	X
6. Contact holes	X		X	
7. First-level metal (missing)	X	X	X	X
8. First-level metal (extra)	X	X	X	X
9. Via holes	X		X	
10. Second-level metal	X		X	
11. Via holes and lead tin	X			
<i>Hot process defects</i>				
1. Voltage-sensitive diffusion shorts		X	X	X
2. SNOS thin oxide shorts		X	X	X
3. MONOS thin oxide shorts			X	X
4. Metal-to-polysilicon shorts			X	X
5. Leakage		X		X
<i>Miscellaneous yield detractors</i>				
1. Ground rules	X		X	
2. Edge limitations (field size)	X			X
3. Circuit yield		X		
4. Fuse success		X		
5. Dicing yield		X		
6. Test probing success		X		
7. Tester repeatability (for multiple tests)		X		
8. Parametric yield				X

learning plans, participation reaches the engineering levels, thus giving everyone a part of the action. The yield plans give management the capability of reviewing progress towards yield objectives. Key problem areas are easily detected and resources can be directed to solve them. Effectiveness of the solution can be measured quantitatively by means of PLY inspection, defect monitors, or test sites.

The success of the yield systems depends on the accuracy of the yield models that are used. Use of the models in the process feedback loop allows them to be constantly evaluated against manufacturing performance and has made it possible to hone them to a peak of perfection.

Appendix A — Yield statistics

The yield and random defect data that have been obtained by the characterization and data handling techniques made one thing abundantly clear: the random defects in IBM manufacturing lines do not behave according to Poisson statistics. Equation (3) may be a nice approximation for the yield, but for yield projections and planning it is incorrect.

It had already been observed in 1964 by B. T. Murphy of the Bell Laboratories that the integrated circuit processes with which he was familiar did not obey Poisson statistics [25]. He proposed the use of mixed or compound Poisson statistics. This observation has subsequently been verified in many semiconductor facilities [26–33].

In PLY inspection, some of the manufacturing lines at IBM eventually used five concentric regions per wafer, while others further divided their concentric regions into quadrants. Each region and quadrant on each wafer had its own defect density for yield calculations. This amounts to using a large number of discrete Poisson distributions. In yield projections, however, using this many parameters was impractical. In the late 1960s, I. F. Chang of IBM solved this difficulty by combining Murphy's mixed Poisson statistics with Dennard's defect sensitivity model. However, it was not until the early seventies that the spread of defect densities was found to be far wider than that assumed by Murphy and Chang. F. Armstrong, O. Paz, and T. R. Lawson, Jr., from IBM East Fishkill, R. S. Hemmert at IBM Manassas, and

C. H. Stapper of IBM Burlington found that the mixing process could, in many cases, be modeled with a gamma distribution [32, 33]. This led to negative binomial or Polya-Eggenberger statistics. At IBM East Fishkill, F. Armstrong and K. Saji studied many different mixing distributions and found that Neyman Type A statistics gave a better fit to some of their data.

Use of mixed Poisson statistics changes the yield expression given in Eq. (3). In the case of negative binomial statistics, the yield becomes

$$Y = Y_0 \left[1 + \left(\sum_i A_{ci} D_i \right) / \alpha \right]^{-\alpha}, \quad (4)$$

and, for Neyman Type A statistics,

$$Y = Y_0 \exp \left\{ -v \left[1 - \exp \left(- \sum_i A_{ci} D_i / v \right) \right] \right\}. \quad (5)$$

The factors α and v in these expressions are parameters resulting from the mixing process. Equations (4) and (5) are equal to (3) when either parameter α or v becomes infinitely large.

Understanding the nature of the defect and fault distributions became of prime importance in the modeling of yield for fault-tolerant memory chips. Models used for this have been described by Stapper, McLaren, and Dreckmann in a previous publication [14].

Appendix B — Process yield

Experience has shown that process yield losses for a particular type of process operation, *e.g.*, diffusion or a photo masking level, are relatively independent of the type of product being manufactured. Hence, the process yield, Y_p , can be estimated by

$$Y_p = \prod_k Y_k^{n_k}, \quad (6)$$

where Y_k is the process yield of the k th type type of process, and n_k is the number of times that process occurs in the manufacturing sequence. By use of this model, process yields for new products with complicated manufacturing processing sequences can be projected from existing process yield data with a high degree of confidence.

Historically, multi-part-number, low-volume products (logic) have had lower process yields in the same line than single-part-number, high-volume products (memory arrays). The difference is due to the increased handling at photo operation to match the correct part-number mask to the right wafers on the logic products. This also leads to greater losses at PAP test and final test, where incorrect masking is typically found.

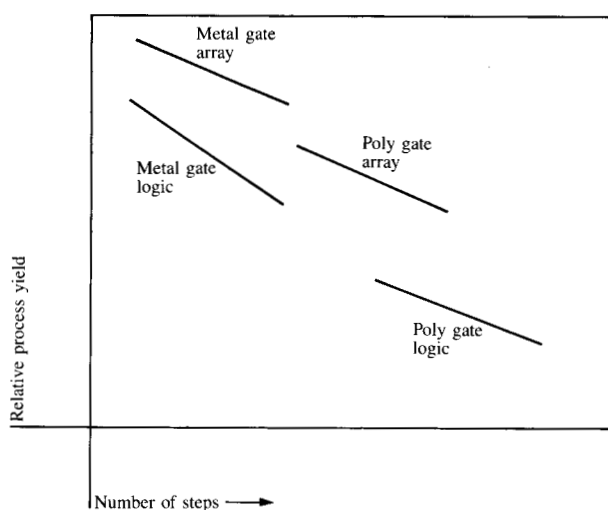


Figure 13 Expected process yield as a function of the increasing number of process steps for various products. Note that the logic products require more handling due to personalization and therefore have lower process yields.

Examples of predictions from this model are given in Fig. 13 for a basic n-channel metal-gate FET process and a polysilicon-gate implanted-junction FET process. Variations in them are projected for both logic and array products. It has been found that process yield learning takes place whenever a new wafer processing line is established. Typically, the learning is relatively rapid during the first year of operation, then gradually it approaches an "ultimate." Manufacturing lines are measured quarterly against such a learning curve as a matter of routine.

It was generally believed, in the 1960s and early 1970s, that a trade-off existed between test and process yields, that high test yields could be achieved by simply scrapping more poor-quality wafers in-line. Conversely, it was believed that higher process yields would occur at the expense of test yields. A system was evolved, over time, whereby process yield targets for each process were established and tracked. Yield losses were analyzed, and the two primary sources of process yield losses were determined to be operator errors and equipment performance. Rigorous operator training and certification programs have been developed and have resulted in a significant improvement in process yields. Likewise, process tool controls have been instituted which require corrective action if the frequency of misprocessing exceeds established limits. The result has been both a reduction in wafer losses due to misprocessing and breakage, and a decrease in defect levels at final test associated with tool and operator performance. Thus, both test and process yields are improved simultaneously.

Acknowledgments

It is impossible to mention everyone who has contributed to yield learning at IBM in this paper. Each yield challenge that has been mastered over the years has brought with it its own team of innovators. Their numbers rank in the thousands. However, we would like to thank two individuals who helped us with this paper: Bob Neudecker, who revised and edited it, and Stan Kirschner, who provided us with valuable ideas, critiques, and also the photographs.

References

1. E. M. Davis, W. E. Harding, R. S. Schwartz, and J. J. Corning, "Solid Logic Technology: Versatile, High-Performance Microelectronics," *IBM J. Res. Develop.* **8**, 102 (1964).
2. P. Frasch and K. H. Saremski, "Feature Size Control in IC Manufacturing," *IBM J. Res. Develop.* **26**, 561 (1982, this issue).
3. G. H. Schwuttke, "Silicon Material Problems in Semiconductor Device Technology," *Microelectron. Reliability* **91**, 397 (1970).
4. H. J. Geipel and W. K. Tice, "Reduction of Leakage by Implantation Gettering in VLSI Circuits," *IBM J. Res. Develop.* **24**, 310 (1980).
5. A. C. Ipri and J. R. Sarace, "Integrated Circuit Process and Design Rule Evaluation," *RCA Rev.* **38**, 323 (1977).
6. D. R. Thomas and R. D. Presson, "An Electrical Photolithographic Alignment Monitor," *Government Microcircuit Application Conference Digest*, 196 (1974).
7. D. P. Kennedy, P. C. Murley, and R. R. O'Brien, "A Statistical Approach to the Design of Diffused Junction Transistors," *IBM J. Res. Develop.* **8**, 482 (1964).
8. C. H. Stapper and P. B. Hwang, "Simulation of FET Device Parameters for LSI Manufacturing," *Proceedings of the Third International Symposium on Silicon Materials, Science and Technology 77-2*, 955 (1977).
9. T. R. Lawson, Jr., "A Prediction of the Photoresist Influence on Integrated Circuit Yield," *Semicond. Prod. Solid State Technol.* **9**, 22 (1966).
10. R. H. Dennard, "Cost Study for Integrated Circuits with Many Logical Decisions per Chip," *Research Report RC-1552*, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, 1966.
11. R. A. Maeder, F. W. Oster, and R. J. Soderman, "Semiconductor and Integrated Circuit Device Yield Modeling," U. S. Patent 3,751,647, 1973.
12. T. Yanagawa, "Yield Degradation of Integrated Circuits Due to Spot Defects," *IEEE Trans. Electron Devices* **ED-19**, 190 (1972).
13. C. H. Stapper, "LSI Yield Modeling and Process Monitoring," *IBM J. Res. Develop.* **20**, 228 (1976).
14. C. H. Stapper, A. N. McLaren, and M. Dreckmann, "Yield Model for Productivity Optimization of VLSI Memory Chips with Redundancy and Partially Good Product," *IBM J. Res. Develop.* **24**, 398 (1980).
15. S. E. Schuster, "Multiple Word/Bit Line Redundancy for Semiconductor Memories," *IEEE J. Solid State Circuits* **SC-13**, 698 (1978).
16. R. P. Cenker, D. G. Clemons, W. R. Huber, J. B. Petrizzi, F. J. Procyk, and G. M. Trout, "A Fault-Tolerant 64K Dynamic RAM," *IEEE ISSCC Digest* **22**, 150 (1979).
17. R. P. Cenker, D. G. Clemons, W. R. Huber, J. B. Petrizzi, F. J. Procyk, and G. M. Trout, "A Fault-Tolerant 64K Dynamic Random-Access Memory," *IEEE Trans. Electron Devices* **ED-26**, 853 (1979).
18. J. F. M. Bindels, J. D. Chlipala, F. H. Fischer, T. F. Mantz, R. G. Nelson, and R. T. Smith, "Cost-Effective Yield Improvement in Fault-Tolerant VLSI Memory," *IEEE ISSCC Digest* **24**, 82 (1981).
19. K. Kokkonen, P. O. Sharp, R. Albers, J. P. Dishaw, F. Louie, and R. J. Smith, "Redundancy Techniques for Fast Static RAMs," *IEEE ISSCC Digest* **24**, 80 (1981).
20. E. A. Reese, D. W. Spaderna, S. T. Flannagan, and F. Tsang, "A 4K $\times$ 8 Dynamic RAM with Self Refresh," *IEEE ISSCC Digest* **24**, 88 (1981).
21. S. S. Eaton, D. Wooten, W. Slemmer, and J. Brady, "A 100ns 64K Dynamic RAM Using Redundancy Techniques," *IEEE ISSCC Digest* **24**, 84 (1981).
22. C. L. Jiang and R. Plachno, "A 32K Static RAM Utilizing a Three-Transistor Cell," *IEEE ISSCC Digest* **24**, 86 (1981).
23. O. Minato, T. Masuhara, T. Sasaki, and Y. Sakai, "H1-CMOSII 4K Static RAM," *IEEE ISSCC Digest* **24**, 14 (1981).
24. E. H. Melan, R. T. Curtis, J. K. Ho, J. G. Koens and G. A. Snyder, "Quality and Reliability Assurance Systems in IBM Semiconductor Manufacturing," *IBM J. Res. Develop.* **26**, 613 (1982, this issue).
25. B. T. Murphy, "Cost Size Optima of Monolithic Integrated Circuits," *Proc. IEEE* **52**, 1537 (1964).
26. W. C. Ansley, "Computation of Integrated Circuit Yields from the Distribution of Slice Yields for the Individual Devices," *IEEE Trans. Electron Devices* **ED-15**, 405 (1968).
27. R. B. Seeds, "Yield, Economics, and Logistic Models for Complex Digital Arrays," *1967 IEEE International Convention Record, Part 6*, 60 (1967).
28. R. M. Warner, Jr., "Applying a Composite Model to the IC Yield Problem," *IEEE J. Solid-State Circuits* **SC-9**, 86 (1974).
29. C. H. Stapper, "On a Composite Model to the IC Yield Problem," *IEEE J. Solid-State Circuits* **SC-10**, 537 (1975).
30. J. Sredni, "Use of Power Transformation to Model the Yield of IC's as a Function of Active Circuit Area," *1975 International Electron Devices Conference Digest*, 123 (1975).
31. J. Bernard, "The IC Yield Problem: A Tentative Analysis of MOS/SOS Circuits," *IEEE Trans. Electron Devices* **ED-25**, 939 (1978).
32. C. H. Stapper, "Defect Density Distribution for LSI Yield Calculations," *IEEE Trans. Electron Devices* **ED-20**, 655 (1973).
33. O. Paz and T. R. Lawson, Jr., "Modification of Poisson Statistics: Modeling Defects Induced by Diffusion," *IEEE J. Solid-State Circuits* **SC-12**, 540 (1977).

Received June 18, 1981; revised March 11, 1982

Paul P. Castrucci IBM General Technology Division, Burlington facility, Essex Junction, Vermont 05452. Mr. Castrucci received his B.S. in physics from Union College, Schenectady, New York, in 1956. In that year he joined IBM as a physicist in the Solid State Device Department at the Poughkeepsie, New York, research laboratory. He is presently manager of silicon gate products manufacturing in Essex Junction. Since joining IBM, he has held numerous engineering and management positions in component development and manufacturing. During his years at IBM, he developed in 1968 the computer industry's initial mono-memory, SP-95, a 16-bit bipolar device, used in IBM's System/360 Model 95; in 1970 the first bipolar monolithic memory to be used in large-scale IBM System/370 Model 145; and in 1971 the first FET mono-memory, a 1024-bit device, used in the IBM System/370 Model 125. He has received his Eighth Plateau of Invention Achievement, five IBM Patent Contribution Awards, and three IBM Outstanding Contribution Awards in 1967, 1969, and 1971. Mr. Castrucci is a member of the Institute of Electrical and Electronics Engineers and the Research Society of America.

Roger A. Maeder *IBM General Technology Division, Burlington facility, Essex Junction, Vermont 05452.* Mr. Maeder is the functional manager of systems and test equipment engineering. He joined IBM in 1965 at Burlington. His past experiences have been in the area of semiconductor yield analysis and process engineering. Mr. Maeder received his A.A.S. in 1961 from Dutchess Community College, Poughkeepsie, New York, a B.S. in electrical engineering from the University of Bridgeport, Connecticut, in 1964, and an M.S. in electrical engineering from the University of Connecticut, Storrs, in 1965. In 1968, Mr. Maeder received an IBM Outstanding Contribution Award for diagnostics and characterization of bipolar memory devices and in 1972 an IBM Outstanding Invention Award for development and implementation of the photo-limited yield model.

William E. Rowe *IBM General Technology Division, Burlington facility, Essex Junction, Vermont 05452.* Mr. Rowe is a senior engineer and project manager for silicon gate technology process development. He joined IBM in 1967 at Essex Junction. While with IBM, he has worked on thin-film process engineering and thin-film manufacturing engineering and has managed FET and bipolar products, characterization, and FET process manufacturing engineering. Mr. Rowe received his B.A. in 1963 and his M.S. in 1965 in physics from the University of Vermont, Burlington.

Charles H. Stapper *IBM General Technology Division, Burlington facility, Essex Junction, Vermont 05452.* Dr. Stapper received his B.S. and M.S. in electrical engineering from the Massachusetts Institute of Technology in 1959 and 1960. After completion of these studies, he joined IBM at the Poughkeepsie, New York, development laboratory, where he worked on magnetic recording and the application of tunnel diodes, magnetic thin films, electron beams, and lasers for digital memories. From 1965 to 1967, he studied at the University of Minnesota on an IBM fellowship. Upon receiving his Ph.D. in 1967, he joined the development laboratory in Essex Junction. His work there included magnetic thin-film array development, magnetic bubble testing and device theory, and bipolar and field effect transistor device theory. He is now a senior engineer in the development laboratory, where he works on mathematical models for yield and reliability management. Dr. Stapper is a member of the Institute of Electrical and Electronics Engineers and Sigma Xi.

Roger A. Verhelst *IBM General Technology Division, Burlington facility, Essex Junction, Vermont 05452.* Dr. Verhelst joined IBM in 1977 at Burlington and is currently manager of the process characterization group for silicon gate products. His earlier assignments included kerf and test site design as well as process characterization for the silicon gate products development pilot line. Prior to joining IBM, Dr. Verhelst was an assistant professor of physics at the University of Maine, Orono, from 1974 to 1977. He received his B.S. in physics and mathematics in 1968 and his Ph.D. in physics in 1973 from Wayne State University, Detroit, Michigan. Dr. Verhelst is a member of the American Physical Society and Sigma Xi.